

# Vertex Cover in $k$ -Uniform Hypergraphs is Hard to Approximate within Factor $k - 3 - \epsilon$

Irit Dinur, Venkatesan Guruswami, Subhash Khot

April 18, 2002

## Abstract

Given a  $k$ -uniform hyper-graph, the  $Ek$ -Vertex-Cover problem is to find the minimum subset of vertices that “hits” every edge. We show that for every integer  $k \geq 5$ , vertex cover in  $k$ -uniform hypergraphs is hard to approximate within factor  $k - 3 - \epsilon$  for an arbitrarily small constant  $\epsilon > 0$ .

This almost matches the upper bound of  $k$  for this problem which is attained by the straightforward greedy approximation algorithm.

The best previously known result was due to Holmerin [Hol02a], showing  $Ek$ -Vertex-Cover hard to approximate to within  $k^{1-\epsilon}$ .

We present two constructions: one with a simple purely-combinatorial analysis, showing  $Ek$ -Vertex-Cover hard to approximate to within  $\Omega(k)$ , followed by a stronger construction that attains the  $k - 3 - \epsilon$  lower bound.

## 1 Introduction

Given a  $k$ -uniform hyper-graph,  $G = (V, E)$  with vertices  $V$  and hyper-edges  $E \subseteq \binom{V}{k} \stackrel{def}{=} \{e \subseteq V \mid |e| = k\}$ , the  $Ek$ -Vertex-Cover problem is the problem of finding the minimum sized subset  $S \subseteq V$  that “hits” every edge in  $G$ , i.e. such that for all  $e \in E$ ,  $e \cap S \neq \emptyset$ . This problem is alternatively called the minimum hitting set with sets of size  $k$ . We show that this problem is NP-hard to approximate to within  $\Omega(k)$  (the constant in the  $\Omega(k)$  is no worse than  $\frac{1}{3}$ ). The result is tight up to the constant in the  $\Omega(k)$  as this problem is approximable to within factor  $k$  by repeatedly selecting one arbitrary hyper-edge, adding all its vertices into the cover and removing all “covered” hyper-edges. The best known algorithm [Hal00] gives only a slight improvement on this greedy algorithm, achieving an approximation factor of  $k - o(1)$ .

## Background

This problem was suggested by Trevisan [Tre01] who initiated a study of bounded degree instances of certain combinatorial problems. There it was shown that this problem is hard to approximate to within  $k^{1/19}$ . Holmerin [Hol02b] showed that E4-Vertex-Cover is NP-hard to approximate to within  $2 - \epsilon$ , and more recently [Hol02a] that  $Ek$ -Vertex-Cover is NP-hard to approximate to within  $k^{1-\epsilon}$ . Goldreich [Gol01] found a simple ‘FGLSS’-type[FGL<sup>+</sup>91] reduction (involving no use of the long-code) to obtain a hardness factor of  $2 - \epsilon$  for  $Ek$ -Vertex-Cover for some constant  $k$ .

## Our Results

We present two constructions: one that is simple attaining a hardness-of-approximation factor of  $\Omega(k)$  (already improving the best previously known result), and one that is stronger attaining a lower bound of  $k - 3 - \epsilon$ .

Our ‘strong’ construction, achieving a hardness factor of  $k - 3 - \epsilon$  involves a novel way of combining ideas from Dinur and Safra’s paper [DS02] and the notion of covering complexity introduced by Guruswami, Håstad and Sudan [GHS00].

Our ‘simple’ construction follows that of [Hol02b] who showed that it is NP-hard to approximate E4-Vertex-Cover to within  $2 - \epsilon$ . Taking a new ‘set-theoretic’ viewpoint on that construction we give a purely-combinatorial proof of Holmerin’s theorem (in contrast to Holmerin’s use of Fourier analysis), relying solely on one new Erdős-Ko-Rado (EKR) type combinatorial lemma (Lemma 2.3) that bounds the maximal size of  $t$ -intersecting families of subsets (i.e. families in

which every pair of subsets intersect on at least  $t$  elements). Taking this new ‘set-theoretic’ viewpoint results in a direct extension of that construction to obtain an  $\Omega(k)$  lower bound.

## Organization of Paper

We begin with some preliminaries, including the starting-point PCP theorem, and some combinatorial lemmas that are used later on in the analysis of the construction. In Section 3 we present our first construction and prove an  $\Omega(k)$  lower bound. In Section 4 we present our stronger construction proving a  $k - 3 - \epsilon$  lower bound.

## Post-Script

Following this work, Dinur, Guruswami, Khot and Regev [DGKR02] were able to show a hardness-of-approximation factor of  $k - 1 - \epsilon$ . The techniques in this paper are quite different and could be of independent interest.

## 2 Preliminaries

For a universe  $R$ , let  $P(R)$  denote its power set, i.e. the family of all subsets of  $R$ . A family  $\mathcal{F} \subseteq P(R)$  is called *monotone* if  $F \in \mathcal{F}, F \subseteq F'$  implies  $F' \in \mathcal{F}$ .

For a “bias parameter”  $0 < p < 1$ , the weight  $\mu_p(F)$  of a set  $F$  is defined as

$$\mu_p(F) \stackrel{\text{def}}{=} p^{|F|}(1-p)^{|R \setminus F|}$$

The weight of a family  $\mathcal{F} \subseteq P(R)$  is defined as

$$\mu_p(\mathcal{F}) \stackrel{\text{def}}{=} \sum_{F \in \mathcal{F}} \mu_p(F).$$

Note that the bias parameter defines a product distribution on  $P(R)$ , where the probability of one subset  $F \in P(R)$  is determined by independently flipping a  $p$ -biased coin to determine the membership of each element of  $R$  in the subset. We denote this distribution by  $\mu_p$ .

### 2.1 Friedgut’s ‘Core’ Theorem

For a family  $\mathcal{F}$ , an element  $\sigma \in R$  and a bias parameter  $p$  we define the “influence of the element on the family” as

$$\text{Influence}_p(\mathcal{F}, \sigma) \stackrel{\text{def}}{=} \Pr_{F \in \mu_p} [\text{exactly one of } F \cup \{\sigma\}, F \setminus \{\sigma\} \text{ is in } \mathcal{F}]$$

The average sensitivity of a family is defined as sum of the influences of all elements.

$$\text{as}_p(\mathcal{F}) \stackrel{\text{def}}{=} \sum_{\sigma \in R} \text{Influence}_p(\mathcal{F}, \sigma)$$

We will use the following theorem that can be obtained by combining Russo’s Theorem and Friedgut’s Theorem. This theorem essentially says that a family of subsets is, in some sense, determined by a ‘core’, see [DS02] for details.

**Theorem 2.1 ([Fri98, ?])** *Let  $p$  be a bias parameter,  $\epsilon, \delta > 0$  be constants and  $\zeta$  be an “accuracy parameter”. Let  $\mathcal{F} \subseteq P(R)$  be a monotone family such that  $\mu_p(\mathcal{F}) \geq \delta$ . Then there exists  $p' \in (p, p + \epsilon)$  and a set  $C \subseteq R$  called the “core” with the following properties :*

- The average sensitivity of the family  $\mathcal{F}$  w.r.t the bias  $p'$  is at most  $\frac{1}{\epsilon}$ , i.e.  $\text{as}_{p'}(\mathcal{F}) \leq \frac{1}{\epsilon}$ .
- The size of  $C$  is a constant that depends only on  $p, \delta, \epsilon, \zeta$ .
- If a family  $\mathcal{H} \subseteq P(R \setminus C)$  is defined as

$$\mathcal{H} \stackrel{\text{def}}{=} \{H \mid H \subseteq R \setminus C, C \cup H \in \mathcal{F}\}$$

Then  $\mu_{p'}(\mathcal{H}) \geq 1 - \zeta$  where  $\mu_{p'}(\mathcal{H})$  is the weight of the family  $\mathcal{H}$  under the  $\mu_{p'}$ -distribution over the universe  $R \setminus C$ .

## 2.2 Intersecting Families of Subsets

In this section we present some combinatorial lemmas regarding intersecting families of subsets, that will be useful later.

**Definition 2.2** For a family of subsets  $\mathcal{F} \subset P(R)$ , let

$$\mathcal{F} \cap \mathcal{F} \stackrel{\text{def}}{=} \{F_1 \cap F_2 \mid F_1, F_2 \in \mathcal{F}\}.$$

**Lemma 2.3 (EKR-Core)** Let  $\epsilon, \delta > 0$ , there exists some  $t = t(\epsilon, \delta) > 0$  such that for every  $\mathcal{F} \subset P(R)$ , if  $\mu_{\frac{1}{2}-\delta}(\mathcal{F}) > \epsilon$ , then there exists some 'core' subset  $C \in \mathcal{F} \cap \mathcal{F}$  with  $|C| \leq t$ .

**Proof Sketch:** A family of subsets is  $t$ -intersecting if for every  $F_1, F_2 \in \mathcal{F}$ ,  $|F_1 \cap F_2| \geq t$ . The idea is that a family cannot be  $t$ -intersecting, for large (but constant)  $t$  and still retain a non-negligible size. Thus, if  $\mu_{\frac{1}{2}-\delta}(\mathcal{F}) > \epsilon$ , there exists some  $t$  for which  $\mathcal{F}$  is not  $t+1$  intersecting, hence  $\mathcal{F}$  has a 'core' subset of size  $t$ .

For the full proof (see Appendix ??), we rely on the complete intersection theorem for finite sets of [AK97] that fully characterizes the maximal  $t$ -intersecting families. ■

**Proposition 2.4** Let  $p > 0$  and let  $\mathcal{F} \subseteq P(R)$ .  $\mu_{p^2}(\mathcal{F} \cap \mathcal{F}) \geq (\mu_p(\mathcal{F}))^2$ .

**Proof:**

$$\Pr_{F \in \mu_{p^2}} [F \in \mathcal{F} \cap \mathcal{F}] = \Pr_{F_1, F_2 \in \mu_p} [F_1 \cap F_2 \in \mathcal{F} \cap \mathcal{F}] \geq \Pr_{F_1 \in \mu_p} [F_1 \in \mathcal{F}] \cdot \Pr_{F_2 \in \mu_p} [F_2 \in \mathcal{F}] = (\mu_p(\mathcal{F}))^2$$

Note that when  $\mathcal{F}$  is defined by exactly one 'minterm', equality holds. ■

**Proposition 2.5** Let  $p > 0$ ,  $\mathcal{F} \subseteq P(R)$ . Let  $\mathcal{F}^k \stackrel{\text{def}}{=} \{F_1 \cap \dots \cap F_k \mid F_i \in \mathcal{F}\}$ . Then,

$$\mu_{p^k}(\mathcal{F}^k) \geq (\mu_p(\mathcal{F}))^k.$$

**Proof:** By induction. ■

## 2.3 Starting Point - PCP

### The Parallel Repetition Theorem

As is the case for many inapproximability results, we begin our reduction from Raz's parallel repetition theorem [Raz98] which is a version of the PCP theorem that is very powerful and convenient to work with. Let  $\Phi = \{\varphi_1, \dots, \varphi_n\}$  be a system of local-constraints over two sets of variables, denoted  $X$  and  $Y$ . Let  $R_X$  denote the range of the  $X$ -variables and  $R_Y$  the range of the  $Y$ -variables<sup>1</sup>. Assume each  $\varphi \in \Phi$  depends on exactly one  $x \in X$  and one  $y \in Y$ , and furthermore, for every value  $a_x \in R_X$  assigned to  $x$  there is exactly one value  $a_y \in R_Y$  to  $y$  such that  $\varphi(a_x, a_y) = \top$ . Therefore, we can write each local constraint  $\varphi \in \Phi$  as a function from  $R_X$  to  $R_Y$ , and use notation  $\varphi_{x \rightarrow y} : R_X \rightarrow R_Y$  (this notation is borrowed from [DS02]). Furthermore, we assume that every  $X$ -variable appears in the same number of local-constraints in  $\Phi$ .

**Theorem 2.6 (PCP Theorem [AS98, ALM<sup>+</sup>98, Raz98])** Let  $\Phi = \{\varphi_1, \dots, \varphi_n\}$  be as above. There exists a universal constant  $\gamma > 0$  such that for every constant  $|R_X|$ , it is NP-hard to distinguish between the following two cases:

- **YES :** There is an assignment  $A : X \cup Y \rightarrow R_X \cup R_Y$  such that all  $\varphi_1, \dots, \varphi_n$  are satisfied by  $A$ , i.e.  $\forall \varphi_{x \rightarrow y} \in \Phi, \varphi_{x \rightarrow y}(A(x)) = A(y)$ .
- **NO :** No assignment can satisfy more than  $\frac{1}{|R_X|^\gamma}$  fraction of  $\Phi$ . ■

<sup>1</sup>Readers familiar with the Raz-verifier may prefer to think concretely of  $R_X = [7^u]$  and  $R_Y = [2^u]$  for some number  $u$  of repetitions.

## Hardness of 4CSP

We define a constraint satisfaction problem (4CSP) which captures a notion related to the notion of covering complexity introduced by [GHS00]. Hardness of this problem is in fact a "reformulation" of a theorem of Holmerin [Hol02b] that will be reproven here (in the next section), as part of our 'simple' construction.

**Definition 2.7** A 4CSP  $\mathcal{L} = (X, \Phi)$  over a domain  $D$  is defined as follows :  $X$  is a set of variables which take values from domain  $D$ . Every  $\phi \in \Phi$  is a constraint on 4 variables. We define YES and NO instances of the CSP as follows.

- **YES :** There exists an assignment  $f : X \mapsto D$  to the variables such that every constraint  $\phi \in \Phi$  is satisfied.
- **NO :** For any subset of variables  $Y \subseteq X$ ,  $|Y| \geq \delta|X|$  and for any  $L$  assignments  $f_1, f_2, \dots, f_L : Y \mapsto D$ , there exists a constraint  $\phi \in \Phi$  such that all the 4 variables of the constraint  $\phi$  are contained in  $Y$  and every assignment  $f_i, 1 \leq i \leq L$  fails to satisfy the constraint  $\phi$ .

We say that  $\phi$  is inside  $Y$  if all the 4 variables of the constraint  $\phi$  are in the set  $Y$ .

**Theorem 2.8** For every integer  $L$  and every constant  $\delta > 0$ , it is NP-hard to distinguish whether an instance  $\mathcal{L}$  of a 4CSP over boolean domain is a YES instance or a NO instance.

**Proof:** Immediate from Holmerin's result (or alternatively, see Theorem 3.1) that for any  $\delta' > 0$ , it is NP-hard to distinguish whether a 4-uniform hypergraph  $G$  is 2-colorable OR  $G$  has no independent set of size  $\delta'|G|$ . Take the Not-all-equal predicate defined by the edges of the hypergraph. ■

**Remark :** The notion of hardness between the YES and NO instances here is closely related to the notion of covering complexity introduced by [GHS00]. The notion of covering complexity requires that in the NO case, no  $L$  assignments satisfy every constraint. We require an even stronger condition that no  $L$  assignments satisfy every constraint inside a set of variables  $Y$  whose size is  $\delta|X|$ .

## 3 The 'Simple' Construction

In this section Our 'simple' construction follows that of [Hol02b] who showed that it is NP-hard to approximate E4-Vertex-Cover to within  $2 - \epsilon$ . Taking a new viewpoint on that construction we give a purely-combinatorial proof of Holmerin's theorem (in contrast to Holmerin's use of Fourier analysis), relying solely on one new Erdős-Ko-Rado (EKR) type combinatorial lemma (Lemma 2.3) that bounds the maximal size of  $t$ -intersecting families of subsets (i.e. families in which every pair of subsets intersect on at least  $t$  elements). We then show a direct extension of that construction to obtain an  $\Omega(k)$  lower bound.

The use of EKR-type bounds in the context of inapproximability results was initiated in [DS02] as part of a more complicated construction and analysis for proving a lower bound for the hardness of approximating vertex-cover on graphs.

The structure of the problem at hand allows a very modular use of EKR-type bounds, and perhaps provides a better intuition as to why they are useful. Since EKR-type bounds are known in many cases to be tight, we believe that similar such bounds may prove fruitful for obtaining improved lower bounds for other approximation problems.

For warmup, let us first prove

**Theorem 3.1** For any  $\delta > 0$ , it is NP-hard to approximate E4-Vertex-Cover to within  $2 - \delta$ .

This result is already known (see [Hol02b]), yet via more complex analysis techniques.

**Proof:** Assume a PCP instance, as given in theorem 2.6, namely a set of local constraints  $\Phi = \{\varphi_1, \dots, \varphi_n\}$  over variables  $X \cup Y$ , whose respective ranges are  $R_X, R_Y$ . For parameters, fix  $t = t(\frac{\delta}{2}, \delta)$ , and take  $|R_X| > (\frac{2t^2}{\delta})^{1/\gamma}$  where  $\gamma > 0$  is the universal constant from Theorem 2.6. From  $\Phi$ , we now construct a 4-uniform hyper-graph whose minimal hitting set has size  $\approx \frac{1}{2}$  or  $\approx 1$  depending on whether  $\Phi$  is satisfiable or not.

We present a construction of a weighted hyper-graph  $G = \langle V, E, \Lambda \rangle$ , which can then be translated into an unweighted hyper-graph via a standard duplication of vertices. The vertex set of  $G$  is

$$V \stackrel{def}{=} X \times P(R_X)$$

namely for each  $x \in X$  we construct a block of vertices denoted  $V[x] = \{x\} \times P(R_X)$  corresponding to all possible subsets of  $R_X$ . The weight of each vertex  $\langle x, F \rangle \in V$  is<sup>2</sup>

$$\Lambda(\langle x, F \rangle) \stackrel{def}{=} \frac{1}{|X|} \cdot \mu_{\frac{1}{2}-\frac{\delta}{4}}(F)$$

The hyper-edges are defined as follows. For every pair of local-constraints  $\varphi_{x_1 \rightarrow y}, \varphi_{x_2 \rightarrow y} \in \Phi$  sharing a mutual variable  $y \in Y$ , we add the hyper edge  $\{\langle x_1, F_1 \rangle, \langle x_1, F'_1 \rangle, \langle x_2, F_2 \rangle, \langle x_2, F'_2 \rangle\}$  only if there is no  $r_1 \in F_1 \cap F'_1$  and  $r_2 \in F_2 \cap F'_2$  such that  $\varphi_{x_1 \rightarrow y}(r_1) = \varphi_{x_2 \rightarrow y}(r_2)$ :

$$E \stackrel{def}{=} \bigcup_{\varphi_{x_1 \rightarrow y}, \varphi_{x_2 \rightarrow y} \in \Phi} \{ \{ \langle x_1, F_1 \rangle, \langle x_1, F'_1 \rangle, \langle x_2, F_2 \rangle, \langle x_2, F'_2 \rangle \} \mid \varphi_{x_1 \rightarrow y}(F_1 \cap F'_1) \cap \varphi_{x_2 \rightarrow y}(F_2 \cap F'_2) = \phi \}$$

where the union is taken over all pairs of local-constraints with a mutual variable  $y$ .

**Lemma 3.2 (Completeness)** *If  $\Phi$  is satisfiable, then  $G$  has a hitting set whose weight is  $\leq \frac{1}{2} + \delta$ .*

**Proof:** Assume a satisfying assignment  $A : X \cup Y \rightarrow R_X \cup R_Y$  for  $\Phi$ . The following set is a hitting set whose weight is  $\frac{1}{2} + \delta$ ,

$$\{ \langle x, F \rangle \in V \mid A(x) \notin F \}$$

For every hyper-edge  $e = \{\langle x_1, F_1 \rangle, \langle x_1, F'_1 \rangle, \langle x_2, F_2 \rangle, \langle x_2, F'_2 \rangle\}$  either  $A(x_1) \notin F_1 \cap F'_1$  or  $A(x_2) \notin F_2 \cap F'_2$ , otherwise since  $A(x_1), A(x_2)$  are restrictions of the same satisfying assignment, they agree on every mutual  $Y$ -variable, so  $\varphi_{x_1 \rightarrow y}(F_1 \cap F'_1) \cap \varphi_{x_2 \rightarrow y}(F_2 \cap F'_2) \neq \phi$ , and  $e$  wouldn't have been a hyper-edge. ■

**Lemma 3.3 (Soundness)** *If  $G$  has a hitting set whose weight is  $\leq 1 - \delta$ , then  $\Phi$  is satisfiable.*

**Proof:** Let  $S \subset V$  be such a hitting set. There must be a set  $X' \subseteq X$  whose fractional size is at least  $\frac{\delta}{2}$ , such that for  $x \in X'$ ,  $\Pr_{v \in \Lambda V} [v \in S \mid v \in V[x]] \leq (1 - \frac{\delta}{2})$ . For each of these blocks, define

$$\mathcal{F}_x = \{ F \in P(R_X) \mid \langle x, F \rangle \notin S \}$$

It follows immediately that  $\forall x \in X', \mu_{\frac{1}{2}-\delta}(\mathcal{F}_x) \geq \frac{\delta}{2}$ . The key observation is that due to lemma 2.3 there exists some 'core' subset  $C \in \mathcal{F}_x \cap \mathcal{F}_x$  whose size is  $|C| \leq t = t(\frac{\delta}{2}, \delta)$ . In other words, there are two subsets which we denote  $F_x^1, F_x^2 \in \mathcal{F}_x$ , such that  $|F_x^1 \cap F_x^2| \leq t$ .

We next translate these 'cores' into an assignment satisfying more than  $\frac{1}{|R_X|^{\gamma}}$  fraction of  $\Phi$ . Let  $x_1, x_2 \in X'$ , and denote their cores respectively by  $C_{x_1} = F_{x_1}^1 \cap F_{x_1}^2, C_{x_2} = F_{x_2}^1 \cap F_{x_2}^2$ . The next observation is that for every  $\varphi_{x_1 \rightarrow y}, \varphi_{x_2 \rightarrow y} \in \Phi$  with  $x_1, x_2 \in X'$ , there always exists some  $r_1 \in C_{x_1}$  and  $r_2 \in C_{x_2}$  such that  $\varphi_{x_1 \rightarrow y}(r_1) = \varphi_{x_2 \rightarrow y}(r_2)$  (i.e.  $\varphi_{x_1 \rightarrow y}(C_{x_1}) \cap \varphi_{x_2 \rightarrow y}(C_{x_2}) \neq \phi$ ). Otherwise the set

$$\{ \langle x_1, F_{x_1}^1 \rangle, \langle x_1, F_{x_1}^2 \rangle, \langle x_2, F_{x_2}^1 \rangle, \langle x_2, F_{x_2}^2 \rangle \}$$

would be a hyper-edge not hit by  $S$ .

Let  $Y' \subseteq Y$  denote the set of all  $Y$  variables that participate in some local-constraint with some  $x \in X'$ ,  $Y' \stackrel{def}{=} \{ y \mid \varphi_{x \rightarrow y} \in \Phi, x \in X' \}$ . Associate each such  $y \in Y'$ , with one arbitrary  $x \in X'$  with  $\varphi_{x \rightarrow y} \in \Phi$ , and let  $C_y \stackrel{def}{=} \varphi_{x \rightarrow y}(C_x) \subset R_Y$ . Now define a random assignment  $A$  by independently selecting for each  $x \in X', y \in Y'$  a random value from  $C_x, C_y$  respectively. Assign the rest of the variables  $(X \setminus X') \cup (Y \setminus Y')$  with any arbitrary value. We complete the proof once we prove

**Proposition 3.4**

$$E_A[\# \{ \varphi_{x \rightarrow y} \text{ is satisfied by } A \}] \geq \frac{\delta}{2t^2} \cdot |\Phi|$$

<sup>2</sup>The requirement that every variable  $x \in X$  appear the same number of times in  $\Phi$  can be avoided by replacing  $\frac{1}{|X|}$  in the weight function with the appropriate proportion of appearances of each  $x$ .

**Proof:** We will show that for any  $x \in X'$ , any  $\varphi_{x \rightarrow y} \in \Phi$  is satisfied by  $A$  with probability  $\geq \frac{1}{t^2}$ ; thus the expected number of local-constraints satisfied by  $A$  is  $\frac{|X'|}{|X|} \cdot \frac{1}{t^2} \geq \frac{\delta}{2t^2} \cdot |\Phi|$  (because every  $x \in X$  appears in the same number of local-constraints). Assume  $C_y = \varphi_{x' \rightarrow y}(C_{x'})$  for some  $x' \in X'$ . Since  $C_x \cap C_{x'} \neq \phi$  implies

$$C_y \cap \varphi_{x \rightarrow y}(C_x) = \varphi_{x \rightarrow y}(C_x) \cap \varphi_{x' \rightarrow y}(C_{x'}) \neq \phi$$

there is at least one value  $a_x \in C_x$  such that  $\varphi_{x \rightarrow y}(a_x) \in C_y$ . Since for every  $x \in X'$ ,  $|C_x| \leq t$ , there is at least  $\frac{1}{t^2}$  probability of having  $\varphi_{x \rightarrow y}(A(x)) = A(y)$ . ■

Thus, there exists some assignment  $A$  that meets the expectation, which means it satisfies  $\geq \frac{\delta}{2t^2} > \frac{1}{|R_X|^7}$  of the local constraints in  $\Phi$ , hence  $\Phi$  is satisfiable. ■

Thus we have proven that distinguishing between the case where the minimal hitting set of  $G$  has size  $\frac{1}{2} + \delta$  to the case of  $1 - \delta$ , enables deciding whether  $\Phi$  is satisfiable or not, hence is NP-hard. ■

### 3.1 $Ek$ -Hitting Set is NP hard to Approximate to Within $k/3 = \Omega(k)$ .

We extend the construction above to work for any constant value of  $k \geq 4$ . We assume wlog that  $k$  is divisible by 4, and for other values of  $k$  we can use the construction for the nearest  $k' = 4m$  and add  $k - k'$  distinct vertices to each edge.

The vertex set for our hyper-graph is the same as in the case for  $k = 4$ , but the weights are different. We set  $p = (\frac{1}{2} - \delta)^{4/k}$ , and

$$\forall v = \langle x, F \rangle \in X \times P(R_X), \quad \Lambda(v) \stackrel{\text{def}}{=} \frac{1}{|X|} \cdot \mu_p(F)$$

The hyper-edges are as follows. For every  $\varphi_{x \rightarrow y}, \varphi_{x' \rightarrow y} \in \Phi$ , and every  $\langle x, F_1 \rangle, \dots, \langle x, F_{\frac{k}{2}} \rangle \in V[x]$  and  $\langle x', F'_1 \rangle, \dots, \langle x', F'_{\frac{k}{2}} \rangle \in V[x']$ , we add the hyper-edge  $\left\{ \langle x, F_1 \rangle, \dots, \langle x, F_{\frac{k}{2}} \rangle, \langle x', F'_1 \rangle, \dots, \langle x', F'_{\frac{k}{2}} \rangle \right\}$  to  $E$  if there is no  $r \in F_1 \cap \dots \cap F_{\frac{k}{2}}$  and  $r' \in F'_1 \cap \dots \cap F'_{\frac{k}{2}}$  with  $\varphi_{x \rightarrow y}(r) = \varphi_{x' \rightarrow y}(r')$ :

$$E = \bigcup_{\varphi_{x \rightarrow y}, \varphi_{x' \rightarrow y} \in \Phi} \left\{ \left\{ \langle x, F_1 \rangle, \dots, \langle x, F_{\frac{k}{2}} \rangle, \langle x', F'_1 \rangle, \dots, \langle x', F'_{\frac{k}{2}} \rangle \right\} \mid \varphi_{x \rightarrow y}(\cap F_i) \cap \varphi_{x' \rightarrow y}(\cap F'_i) = \phi \right\}$$

As in the case of  $k = 4$ ,

**Lemma 3.5 (Completeness)** *If  $\Phi$  is satisfiable, then  $G$  has a hitting set whose weight is  $\leq 1 - p = \frac{O(1)}{k}$ .*

**Proof:** Again we take  $A$  to be a satisfying assignment, and set  $S = \bigcup_{x \in X} \{ \langle x, F \rangle \mid F \not\subseteq A(x) \}$ . The weight of  $S$  is  $1 - p$ . ■

The proof of soundness is also quite similar, with one minor twist.

**Lemma 3.6 (Soundness)** *If  $G$  has a hitting set whose weight is  $\leq 1 - \delta$ , then  $\Phi$  is satisfiable.*

**Proof:** Let  $S \subset V$  be such a hitting set. Again we consider the set  $X' \subseteq X$  for which  $\forall x \in X', \Pr_{v \in \Lambda V} [v \in S \mid v \in V[x]] \leq (1 - \frac{\delta}{2})$ . Again  $\frac{|X'|}{|X|} \geq \frac{\delta}{2}$ . For each  $x \in X'$  let

$$\mathcal{F}_x = \{ F \in P(R_X) \mid \langle x, F \rangle \notin S \}$$

We cannot immediately apply lemma 2.3 to find 'weak-cores' for each  $\mathcal{F}_x$  because lemma 2.3 works only when  $p < \frac{1}{2} - \delta$ . Thus, we must first consider the family  $\mathcal{F}'_x = \{ F_1 \cap \dots \cap F_{k/4} \mid F_i \in \mathcal{F}_x \}$  whose size is guaranteed to be, by proposition 2.5,

$$\mu_{\frac{1}{2}-\delta}(\mathcal{F}'_x) = \mu_{p^{k/4}}(\mathcal{F}'_x) \geq (\mu_p(\mathcal{F}_x))^{k/4} = \delta^{k/4}$$

Now we can deduce the existence of a 'core' subset  $C_x \subset R_X$ ,  $|C_x| \leq t = t(\delta^{k/4}, \delta)$ , such that there are some  $F_x^1, \dots, F_x^{k/4} \in \mathcal{F}_x$  so that  $C_x = F_x^1 \cap \dots \cap F_x^{k/4}$ . Again, observe that for every  $\varphi_{x \rightarrow y}, \varphi_{x' \rightarrow y} \in \Phi$ ,  $\varphi_{x \rightarrow y}(C_x) \cap \varphi_{x' \rightarrow y}(C_{x'}) \neq \phi$ . From here we can proceed as in the proof of lemma 3.3 above to define a random assignment that expectedly satisfies  $\frac{\delta}{2t^2}$  of  $\Phi$ . ■

All in all, we have proven

**Theorem 3.7** *It is NP-hard to approximate  $Ek$ -Vertex-Cover to within  $k/3$ .*

**Proof:** For a  $k$ -uniform hyper-graph, we have proven in Lemmas 3.5,3.6 that it is NP-hard to distinguish between a vertex-cover of size  $1 - \delta$  and  $1 - p = 1 - (\frac{1}{2} - \delta)^{4/k} < \frac{3}{k}$ . The last inequality follows from

$$(1 - \frac{3}{k})^{k/4} < \frac{1}{e^{3/4}} < \frac{1}{2} - \delta$$

which implies that there is no  $\frac{k}{3}$  approximation algorithm for vertex-cover on  $k$ -uniform hyper-graphs, unless P=NP. ■

## 4 The ‘Stronger’ Construction

Let  $\mathcal{B}$  be the set of all  $l$ -tuples of variables of an instance  $\mathcal{L}$  of a 4CSP given by Theorem 2.8. That is

$$\mathcal{B} \stackrel{\text{def}}{=} \{(x_1, x_2, \dots, x_l) \mid x_i \in X\}$$

An  $l$ -tuple  $B \in \mathcal{B}$  will be called a “block”. Let  $R$  be the set of all possible “block assignments”, i.e.  $R \stackrel{\text{def}}{=} D^l$  is the set of all strings of length  $l$  over the domain  $D$ . Let  $P(R)$  denote the family of all subsets of  $R$ , i.e.

$$P(R) \stackrel{\text{def}}{=} \{F \mid F \subseteq R\}$$

The vertex set  $V$  of the hypergraph is defined to be

$$V \stackrel{\text{def}}{=} \mathcal{B} \times P(R) = \{(B, F) \mid B \in \mathcal{B}, F \in P(R)\}$$

The vertices will have weights. Let  $p = 1 - \frac{1}{k-3} - \epsilon$  be the “bias parameter”. The weight of a vertex  $(B, F)$  is  $\mu_p(F)$  where

$$\mu_p(F) \stackrel{\text{def}}{=} p^{|F|} (1-p)^{|R \setminus F|}$$

To motivate the way we define the edges of the hypergraph, assume that  $f : X \mapsto D$  is an assignment that satisfies every constraint. Let  $f[B]$  denote the restriction of this assignment to block  $B$ . Thus  $f[B] \in R$ . The edges of the hypergraph will be defined in such a way that the set of vertices  $\mathcal{I}$

$$\mathcal{I} = \{(B, F) \mid B \in \mathcal{B}, f[B] \in F\} \tag{1}$$

is an independent set.

**Definition 4.1** *We say that 4 blocks  $(B_1, B_2, B_3, B_4)$  are “overlapping” if they agree on some  $l-1$  coordinates and the 4 variables on the remaining coordinate form a constraint in the 4CSP. More precisely, there exist variables  $x_1, x_2, \dots, x_{l-1}$  and  $y_1, y_2, y_3, y_4$  and an index  $t, 1 \leq t \leq l$  such that*

1.  $B_i = (x_1, x_2, \dots, x_{t-1}, y_i, x_t, x_{t+1}, \dots, x_{l-1})$  for  $i = 1, 2, 3, 4$
2. There is a constraint  $\phi \in \Phi$  on the variables  $(y_1, y_2, y_3, y_4)$ .

*Note that the tuple  $(\{x_j\}_{j=1}^{t-1}, \{y_i\}_{i=1}^4, t, \phi)$  completely characterizes the overlapping blocks.*

For a block  $B = (z_1, z_2, \dots, z_l)$  and a block assignment  $\sigma \in R$ , let  $\sigma(z_j)$  denote the value assigned by  $\sigma$  to the variable  $z_j$ , which is just the  $j^{\text{th}}$  coordinate of  $\sigma$ . For  $1 \leq j \leq l$ , let  $\pi_j : D^l \mapsto D^{l-1}$  be the projection operator that maps a string of length  $l$  to its substring of length  $l-1$  obtained by dropping the  $j^{\text{th}}$  coordinate.

**Definition 4.2** *For any overlapping blocks  $(B_1, B_2, B_3, B_4)$ , characterized by  $(\{x_j\}_{j=1}^{t-1}, \{y_i\}_{i=1}^4, t, \phi)$ , and block assignments  $\sigma^{(i)}$  to the blocks  $B_i$ s, we say that these block assignments are consistent if*

1.  $\pi_t(\sigma^{(1)}) = \pi_t(\sigma^{(2)}) = \pi_t(\sigma^{(3)}) = \pi_t(\sigma^{(4)})$

2. The values  $\sigma^{(1)}(y_1), \sigma^{(2)}(y_2), \sigma^{(3)}(y_3), \sigma^{(4)}(y_4)$  satisfy the constraint  $\phi$ .

In short, the first condition says that the assignments  $\sigma^{(i)}$  must “project” down to a common assignment to  $(l - 1)$  coordinates and the second condition says that the 4 values on the remaining coordinate must satisfy the constraint  $\phi$ .

Note that if  $f : X \mapsto D$  is an assignment that satisfies every constraint, and  $f[B]$  is the restriction of this assignment to a block  $B$ , then for any overlapping blocks  $(B_1, B_2, B_3, B_4)$ , the block assignments  $f[B_1], f[B_2], f[B_3], f[B_4]$  are consistent.

**Definition 4.3** For overlapping blocks  $(B_1, B_2, B_3, B_4)$ , and  $k$  sets  $F_1, F_2, \dots, F_{k-3}, F^{(2)}, F^{(3)}, F^{(4)} \subseteq R$ , we say that these  $k$  sets are consistent if there exist block assignments  $\sigma^{(i)}$  for block  $B_i$  such that

1.  $\sigma^{(1)} \in F_1 \cap F_2 \cap \dots \cap F_{k-3}$
2.  $\sigma^{(i)} \in F^{(i)}$  for  $i = 2, 3, 4$ .
3. The assignments  $\sigma^{(i)}$  are consistent as per Definition 4.2.

**Remark :** Whenever we talk about consistency between sets  $F_1, F_2, \dots, F_{k-3}, F^{(2)}, F^{(3)}, F^{(4)}$ , we have in mind a specific set of overlapping blocks  $(B_1, B_2, B_3, B_4)$  which we will be clear from the context.

Now we are ready to define edges of the hypergraph. For overlapping blocks  $(B_1, B_2, B_3, B_4)$ , and sets  $F_1, F_2, \dots, F_{k-3}, F^{(2)}, F^{(3)}, F^{(4)}$  which are **not consistent**, we define

$$\{(B_1, F_j) | j = 1, 2, \dots, k-3\} \cup \{(B_i, F^{(i)}) | i = 2, 3, 4\}$$

to be an edge of the hypergraph. Thus every edge contains exactly  $k$  vertices, i.e. this is a  $k$ -uniform hypergraph.

Lets verify that this way of defining edges makes sense. Suppose  $f : X \mapsto D$  is an assignment that satisfies every constraint. We will show that the set  $\mathcal{I}$  (see Equation (1)) is an independent set. As observed before, for any overlapping blocks  $(B_1, B_2, B_3, B_4)$ , the block assignments  $f[B_i]$  are consistent. Let

$$\{(B_1, F_j) | j = 1, 2, \dots, k-3\} \cup \{(B_i, F^{(i)}) | i = 2, 3, 4\}$$

be any  $k$  vertices in the set  $\mathcal{I}$ . We will show that the sets

$$F_1, F_2, \dots, F_{k-3}, F^{(2)}, F^{(3)}, F^{(4)}$$

are consistent and hence these  $k$  vertices **cannot** form an edge, thus proving that  $\mathcal{I}$  is indeed an independent set. By definition of the set  $\mathcal{I}$ , we have  $f[B_1] \in F_1 \cap F_2 \cap \dots \cap F_{k-3}$  and  $f[B_i] \in F^{(i)}$  for  $i = 2, 3, 4$ . Since the assignments  $f[B_i]$  are consistent, taking  $\sigma^{(i)} = f[B_i]$  in Definition 4.3 proves the claim.

## 4.1 Completeness

We will show that if there is a global assignment  $f : X \mapsto D$  that satisfies every constraint, then the hypergraphs has a “large” independent set. As observed in the last section, the set

$$\mathcal{I} = \{(B, F) \mid B \in \mathcal{B}, f[B] \in F\}$$

is an independent set. The weight of this set is

$$\sum_{B \in \mathcal{B}} \sum_{F: F \subseteq R, f[B] \in F} \mu_p(F) = \sum_{B \in \mathcal{B}} p = p|\mathcal{B}|$$

Thus in the completeness case, there exists an independent set of size  $p|\mathcal{B}|$ .



## 4.2 Soundness

We will show that if the 4CSP instance  $\mathcal{L}$  is a NO instance, then the hypergraph we constructed has no independent set of size  $\delta|\mathcal{B}|$ . Thus we will have a gap of  $(p|\mathcal{B}|, \delta|\mathcal{B}|)$  in the size of the independent set which corresponds to a gap  $((1-p)|\mathcal{B}|, (1-\delta)|\mathcal{B}|)$  in the size of the vertex cover. This is a factor  $\frac{1-\delta}{1-p} = \frac{1-\delta}{\frac{1}{k-3}+\epsilon} = k-3-\epsilon'$  gap where  $\epsilon' \rightarrow 0$  as  $\epsilon, \delta \rightarrow 0$ .

Assume on the contrary that the hypergraph has an independent set of size  $\delta|\mathcal{B}|$ . Call this independent set  $\mathcal{I}$ . We will construct a collection of  $|\mathcal{D}|^{t-1}$  assignments to a set of variables  $Y, |Y| \geq \delta|X|/2$  in the 4CSP such that every constraint inside  $Y$  is satisfied by some assignment.

For every block  $B$ , define

$$\mathcal{F}[B] \stackrel{\text{def}}{=} \{F \mid F \subseteq R, (B, F) \in \mathcal{I}\}$$

A simple averaging argument shows that for at least  $\delta/2$  fraction of the blocks  $B$ , we have  $\mu_p(\mathcal{F}[B]) \geq \delta/2$ . Defining

$$\mathcal{B}' \stackrel{\text{def}}{=} \{B \mid B \in \mathcal{B}, \mu_p(\mathcal{F}[B]) \geq \delta/2\}$$

we have  $|\mathcal{B}'| \geq \delta|\mathcal{B}|/2$ .

**Lemma 4.4** *For every  $B \in \mathcal{B}$ , the family  $\mathcal{F}[B]$  can be assumed to be a monotone family of subsets of  $R$ .*

**Proof:** The way we define the edges of the hypergraph, it is easy to see that if  $(B, F)$  is a vertex of an independent set then we can also add  $(B, F')$  to the independent set provided  $F \subseteq F'$ . Thus when the independent set is maximal, every family  $\mathcal{F}[B]$  is monotone.  $\blacksquare$

Using this lemma, for every  $B \in \mathcal{B}'$ , the family  $\mathcal{F}[B]$  is a monotone family with  $\mu_p(\mathcal{F}[B]) \geq \delta/2$ . Let  $\zeta > 0$  be a sufficiently small “accuracy” parameter which will be fixed later. Applying Theorem 2.1, we get

**Lemma 4.5** *For every block  $B \in \mathcal{B}'$ , there exists a real number  $p[B] \in (p, p + \frac{\epsilon}{2})$  and a set  $C[B] \subseteq R$  called the “core” with the following properties :*

- $\text{as}_{p[B]}(\mathcal{F}[B]) \leq \frac{2}{\epsilon}$ .
- The size of  $C[B]$  is at most  $\Delta_0$  which is a constant depending only on  $k, \epsilon, \zeta, \delta$ .
- Let  $\mathcal{H}[B] \subseteq R \setminus C[B]$  defined as

$$\mathcal{H}[B] \stackrel{\text{def}}{=} \{H \mid H \subseteq R \setminus C[B], C[B] \cup H \in \mathcal{F}[B]\}$$

Then we have  $\mu_{p[B]}(\mathcal{H}[B]) \geq 1 - \zeta$ , where the weight of the family  $\mathcal{H}[B]$  is measured w.r.t. the  $p[B]$ -distribution on the universe  $R \setminus C[B]$ .

Let  $\eta > 0$  be a threshold parameter which will be chosen later. For every  $B \in \mathcal{B}'$ , we identify a set of elements  $\text{Infl}[B] \subseteq R$  that have significant influence on the family  $\mathcal{F}[B]$ , i.e.

$$\text{Infl}[B] = \{\sigma \in R \mid \text{Influence}_{p[B]}(\mathcal{F}[B], \sigma) \geq \eta\}$$

Since  $\mathcal{F}[B]$  has average sensitivity at most  $\frac{2}{\epsilon}$  and the average sensitivity is simply the sum of influences of all the elements, it follows that the size of  $\text{Infl}[B]$  is at most  $\frac{2}{\eta\epsilon}$  which is a constant. Finally define the “extended core”  $\text{Ecore}[B]$  as

$$\text{Ecore}[B] \stackrel{\text{def}}{=} C[B] \cup \text{Infl}[B]$$

Clearly, the extended core has size at most  $\Delta = \Delta_0 + \frac{2}{\eta\epsilon}$ .

### 4.3 The Preservation Property

Given two block assignments  $\sigma, \sigma'$ , and a projection  $\pi_j : D^l \mapsto D^{l-1}, 1 \leq j \leq k$ , we say that the two assignments are “preserved” if  $\pi_j(\sigma) \neq \pi_j(\sigma')$ . Since  $\sigma, \sigma'$  differ in at least one coordinate, they will be preserved with probability  $1 - \frac{1}{l}$  when a projection  $\pi_j, 1 \leq j \leq l$  is picked at random.

For a block  $B \in \mathcal{B}'$ , say that its extended core is preserved under projection  $\pi_j$  if every pair of elements in the extended core is preserved. In other words, the projection operator is one-to-one on the extended core.

The extended core has size at most  $\Delta$ . Choosing  $l = \Delta^2$ , the probability that the extended core is preserved under a random projection  $\pi_j$ , is at least  $1 - \frac{\binom{\Delta}{2}}{l} \geq \frac{1}{2}$ . Hence there exists an index  $j_0, 1 \leq j_0 \leq l$ , such that for at least half of the blocks in  $\mathcal{B}'$ , their extended core is preserved. Assume w.l.o.g. that  $j_0 = l$  and  $\pi = \pi_l$  denote the projection operator which acts simply by dropping the last coordinate.

$$\mathcal{B}'' \stackrel{\text{def}}{=} \{B \mid B \in \mathcal{B}', \text{Ecore}[B] \text{ is preserved by } \pi\}$$

As noted,  $|\mathcal{B}''| \geq |\mathcal{B}'|/2 \geq \delta|\mathcal{B}|/4$ . A simple averaging argument shows that we can fix variables  $x_1, x_2, \dots, x_{l-1} \in X$  such that for at least  $\delta/4$  fraction of variables  $y$ , we have  $(x_1, x_2, \dots, x_{l-1}, y) \in \mathcal{B}''$ . Define

$$Y \stackrel{\text{def}}{=} \{y \mid y \in X, (x_1, x_2, \dots, x_{l-1}, y) \in \mathcal{B}''\}$$

Thus we have  $|Y| \geq \delta|X|/4$ . Denote by  $B_y$  the block  $(x_1, x_2, \dots, x_{l-1}, y)$ .

### 4.4 Defining Assignments

Now we are ready to define assignments to the variables in set  $Y$  so that every constraint inside  $Y$  is satisfied by some assignment. There will be one assignment  $f_\tau : Y \mapsto D$  for every  $\tau \in D^{l-1}$ . For  $\tau \in D^{l-1}$  and  $\alpha \in D$ , let  $\tau\alpha \in R = D^l$  be the concatenated string.

The assignment  $f_\tau : Y \mapsto D$  is defined as

$$f_\tau(y) = \begin{cases} \alpha & \text{if } \exists \alpha \in D \text{ s.t. } \tau\alpha \in \text{Ecore}[B_y] \\ \text{undefined} & \text{otherwise} \end{cases}$$

There are two things to note here. Firstly, since the extended core is preserved, there exists at most one  $\alpha \in D$  such that  $\tau\alpha \in \text{Ecore}[B_y]$ . Thus the definition of  $f_\tau$  is unambiguous. Secondly, though the assignment  $f_\tau$  is undefined for some (or even all) of the variables in  $Y$ , we will still show that for every constraint  $\phi$  inside  $Y$ , there exists an assignment  $f_\tau$  such that it satisfies the constraint  $\phi$ . We prove this in the next section.

### 4.5 The Main Proof

In this section, we will show that for every constraint  $\phi$  inside the set of variables  $Y$ , there exists an assignment  $f_\tau$  that satisfies this constraint. Let  $\phi$  be a constraint on the variables  $\{y_1, y_2, y_3, y_4\}$  and consider the blocks

$$B_i = B_{y_i} = (x_1, x_2, \dots, x_{l-1}, y_i)$$

Clearly, the blocks  $(B_1, B_2, B_3, B_4)$  are overlapping. We prove our claim in several steps.

**Lemma 4.6** *There exist sets  $F'_1, F'_2, \dots, F'_{k-3} \subseteq R \setminus C[B_1]$  such that*

- $\bigcap_{j=1}^{k-3} F'_j = \phi$
- $F'_j \stackrel{\text{def}}{=} C[B_1] \cup F'_j \in \mathcal{F}[B_1] \text{ for } 1 \leq j \leq k-3$ .

*In particular,  $\bigcap_{j=1}^{k-3} F_j = C[B_1]$ .*

**Proof:** From Lemma 4.5, the weight of the family  $\mathcal{H}[B_1]$  w.r.t. the bias parameter  $p[B]$  is at least  $1 - \zeta$ . Noting that  $p[B] \leq 1 - \frac{1}{k-3} - \frac{\epsilon}{2}$  and applying Lemma A.3, there exist sets  $F'_j \subseteq R \setminus C[B_1], 1 \leq j \leq k-3$  whose intersection is empty. By definition of the family  $\mathcal{H}[B_1]$ , the sets  $F_j \stackrel{\text{def}}{=} C[B_1] \cup F'_j \in \mathcal{F}[B_1]$ . ■

Define

$$S \stackrel{\text{def}}{=} \{\sigma \in R \mid \text{there exists } \sigma' \in C[B_1] \text{ such that } \pi(\sigma) = \pi(\sigma')\}$$

That is,  $S$  is the set of all strings which share a common prefix of length  $l - 1$  with *some* string in  $C[B_1]$ . Clearly  $|S| = |D| \cdot |C[B_1]| \leq |D| \cdot \Delta_0$ . For  $i = 2, 3, 4$  define

$$T_i \stackrel{\text{def}}{=} S \setminus \text{Ecore}[B_i]$$

Thus  $T_i$  is a set of size at most  $|D| \cdot \Delta_0$ . By definition of the extended core (at the end of the Section 4.2), all elements of the set  $T_i$  have influence at most  $\eta$  on the family  $\mathcal{F}[B_i]$  w.r.t. bias  $p[B_i]$ . Applying Lemma A.1, if  $\eta$  is small enough, there exists a set  $F^{(i)} \in \mathcal{F}[B_i]$  such that  $F^{(i)} \cap T_i = \phi$ .

Now consider the following vertices of the hypergraph :

$$\{(B_1, F_j) \mid 1 \leq j \leq k - 3\} \cup \{(B_i, F^{(i)}) \mid i = 2, 3, 4\}$$

These vertices are in the independent set  $\mathcal{I}$ . Hence there **can't** be an edge on these vertices. Therefore the sets  $F_1, \dots, F_j, F^{(2)}, F^{(3)}, F^{(4)}$  are consistent (see Definition 4.3). This means that there exist block assignments  $\sigma^{(i)} \in R$  such that

- $\sigma^{(1)} \in \bigcap_{j=1}^{k-3} F_j = C[B_1]$  (by Lemma 4.6).
- For  $i = 2, 3, 4$ ,  $\sigma^{(i)} \in F^{(i)}$ .
- The block assignments  $\sigma^{(i)}$  have the same prefix of length  $l - 1$ , i.e. there is a string  $\tau \in D^{l-1}$ , and values  $\alpha_i \in D$  such that  $\sigma^{(i)} = \tau \alpha_i$ .
- The values  $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$  satisfy the constraint  $\phi$ .

**Lemma 4.7**  $\sigma^{(i)} \in \text{Ecore}[B_i]$  for  $i = 1, 2, 3, 4$ .

**Proof:** We have  $\sigma^{(1)} \in C[B_1] \subseteq \text{Ecore}[B_1]$ . Now consider  $i = 2, 3, 4$ . Since  $\sigma^{(i)}$  has the same  $(l - 1)$ -prefix with  $\sigma^{(1)}$ , by definition of the set  $S$ ,  $\sigma^{(i)} \in S$ . Also  $\sigma^{(i)} \in F^{(i)}$  and  $F^{(i)} \cap T_i = \phi$ . Therefore  $\sigma^{(i)} \in \text{Ecore}[B_i]$ . ■

From this lemma, and the way the assignment  $f_\tau$  is defined, we have  $f_\tau(y_i) = \alpha_i$  for  $i = 1, 2, 3, 4$ . Since the values  $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$  satisfy the constraint  $\phi$ , it follows that the assignment  $f_\tau$  satisfies the constraint  $\phi$ . This finishes the proof.

## 5 Future Work

The vertex cover in every  $k$ -uniform hyper-graph can be approximated to within factor  $k - o(1)$ , [Hal00]. An obvious open problem is that of obtaining a matching lower bound of  $k - \epsilon$ , i.e. proving NP-hardness for approximating  $Ek$ -Vertex-Cover to within  $k - \epsilon$  for any  $\epsilon > 0$ .

Another possible direction is to extend these results for larger values of  $k$ . The largest plausible value of  $k$  is  $\ln n$  since the greedy set-cover algorithm can always be used to achieve a  $\ln n + 1$  approximation on any hyper-graph. We sum this up with the following conjecture:

**Conjecture 5.1** *It is NP-hard to approximate  $Ek$ -Vertex-Cover to within  $k \cdot (1 - \epsilon)$  for any  $k \leq \ln n$  and any  $\epsilon > 0$ .*

## References

- [AK97] Rudolf Ahlswede and Levon H. Khachatrian. The complete intersection theorem for systems of finite sets. *European J. Combin.*, 18(2):125–136, 1997.
- [ALM<sup>+</sup>98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, May 1998.

- [AS98] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: a new characterization of NP. *Journal of the ACM*, 45(1):70–122, January 1998.
- [DGKR02] I. Dinur, V. Guruswami, S. Khot, and O. Regev. Paper yet to be written, 2002.
- [DS98] I. Dinur and S. Safra. Monotone-minimum-satisfying assignment is NP-hard for almost polynomial factors. Manuscript, 1998.
- [DS02] I. Dinur and S. Safra. On the importance of being biased. In *Proc. 34th ACM Symp. on Theory of Computing*, 2002.
- [FGL<sup>+</sup>91] U. Feige, S. Goldwasser, L. Lovász, S. Safra, and M. Szegedy. Approximating clique is almost NP-complete. In *Proc. 32nd IEEE Symp. on Foundations of Computer Science*, pages 2–12, 1991.
- [Fri98] Ehud Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–35, 1998.
- [GHS00] V. Guruswami, J. Hastad, and M. Sudan. Hardness of approximate hypergraph coloring. In *Proc. 41st IEEE Symp. on Foundations of Computer Science*, pages 149–158. IEEE Computer Society Press, 2000.
- [Gol01] O. Goldreich. Using the FGLSS-reduction to prove inapproximability results for minimum vertex cover in hypergraphs. ECCC Technical Report TR01-102, 2001.
- [Hal00] Eran Halperin. Improved approximation algorithms for the vertex cover problem in graphs and hypergraphs. In *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 329–337, N.Y., January 9–11 2000. ACM Press.
- [Hol02a] J. Holmerin. Improved inapproximability results for vertex cover on k-regular hyper-graphs. 2002.
- [Hol02b] J. Holmerin. Vertex cover on 4-regular hyper-graphs is hard to approximate within  $2 - \epsilon$ . In *Proc. 34th ACM Symp. on Theory of Computing*, 2002.
- [Raz98] Ran Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, June 1998.
- [Tre01] L. Trevisan. Non-approximability results for optimization problems on bounded degree instances. In *Proc. 33rd ACM Symp. on Theory of Computing*, 2001.

## A Some useful lemmas

The following lemma can be found in [DS98].

**Lemma A.1** *Let  $\mathcal{F} \subseteq P(R)$  be a monotone family. Let  $T$  be a set of elements such that for every element  $\sigma \in T$ ,  $\text{Influence}(\mathcal{F}, \sigma, p) < \eta$ . Assume  $\eta$  is small enough so that*

$$|T| \cdot \eta \cdot p^{-|T|} < \mu_p(\mathcal{F})$$

*Then there exists a set  $F \in \mathcal{F}$  such that  $F \cap T = \emptyset$ .*

### A.1 $k$ -wise Intersecting Families

We will use the following theorem of Frankl.

**Theorem A.2** *Let  $\mathcal{F} \subseteq P(R)$  where  $|R| = n$  and every set in the family  $\mathcal{F}$  has size  $m$ . Assume that every  $k$  sets in the family have nonempty intersection and  $n > mk/(k-1)$ . Then*

$$|\mathcal{F}| \leq \binom{n-1}{m-1}$$

Note that a family of sets of size  $m$  containing one fixed element has size  $\binom{n-1}{m-1}$ . We will use the above theorem to prove :

**Lemma A.3** *Let  $\epsilon > 0$  be an arbitrarily small constant,  $k \geq 2$  an integer and  $p = 1 - \frac{1}{k} - \epsilon$ . Let  $\mathcal{F} \subseteq P(R)$  be a family such that every  $k$  sets in this family have a nonempty intersection. Then*

$$\mu_p(\mathcal{F}) < p + \epsilon$$

*provided the universe  $R$  is sufficiently large.*

**Proof:** Let  $n = |R|$  be the size of the universe. Partition the family  $\mathcal{F}$  according to different set-sizes.

$$\mathcal{F}_i \stackrel{\text{def}}{=} \{F \mid F \in \mathcal{F}, |F| = i\}$$

With the bias parameter  $p$ , the total weight of all sets of size more than  $(p + \frac{\epsilon}{2})n$  is at most  $\frac{\epsilon}{2}$  when the universe is large enough. Hence

$$\mu_p(\mathcal{F}) \leq \frac{\epsilon}{2} + \sum_{m \leq (p + \frac{\epsilon}{2})n} \mu_p(\mathcal{F}_m)$$

For  $m \leq (p + \frac{\epsilon}{2})n$ , we have  $n > mk/(k-1)$ . Since every  $k$  sets in the family  $\mathcal{F}_m$  have a nonempty intersection, applying Frankl's Theorem, we get

$$|\mathcal{F}_m| \leq \binom{n-1}{m-1}$$

Noting that every set in  $\mathcal{F}_m$  has weight  $p^m(1-p)^{n-m}$  we have

$$\begin{aligned} \mu_p(\mathcal{F}) &\leq \frac{\epsilon}{2} + \sum_{m \leq (p + \frac{\epsilon}{2})n} \binom{n-1}{m-1} p^m (1-p)^{n-m} \\ &\leq \frac{\epsilon}{2} + p \left( \sum_m \binom{n-1}{m-1} p^{m-1} (1-p)^{(n-1)-(m-1)} \right) \\ &= \frac{\epsilon}{2} + p \end{aligned}$$

■

## A.2 EKR core

**Lemma 2.3 (Core)** *Let  $\epsilon, \delta > 0$ , there exists some  $t = t(\epsilon, \delta) > 0$  such that for every finite  $R$  and  $\mathcal{F} \subset P(R)$ , if  $\mu_{\frac{1}{2}-\delta}(\mathcal{F}) > \epsilon$ , then there exists some 'core' subset  $C \in \mathcal{F} \cap \mathcal{F}$  with  $|C| \leq t$ .*

**Proof:** We begin by stating a continuous variant of the complete intersection theorem of Ahlswede and Khachatrian. This was already proven in [DS02] for  $t = 2$  and the extension for larger  $t$  is straightforward.

Define for every  $i \geq 0$ ,  $t > 0$  and  $n \geq 2i + t$ ,

$$\mathcal{A}_{i,t}^n \stackrel{def}{=} \{F \in P([n]) \mid F \cap [1, t + 2i] \geq t + i\}.$$

Clearly, for any  $n' > n \geq 2i + t$ ,  $\mu_p(\mathcal{A}_{i,t}^{n'}) = \mu_p(\mathcal{A}_{i,t}^n)$ . Denoting  $\binom{[n]}{k} \stackrel{def}{=} \{F \subset [n] \mid |F| = k\}$ , the complete intersection theorem of Ahlswede and Khachatrian states that

**Theorem A.4 ([AK97])** *Let  $\mathcal{F} \subseteq \binom{[n]}{k}$  be  $t$ -intersecting. Then,*

$$|\mathcal{F}| \leq \max_{0 \leq i \leq \frac{n-t}{2}} \left| \mathcal{A}_{i,t}^n \cap \binom{[n]}{k} \right|$$

■

The following lemma is a continuous variant of the above theorem,

**Lemma A.5** *[[DS02] Let  $\mathcal{F} \subset P([n])$  be  $t$ -intersecting. For any  $p < \frac{1}{2}$ ,*

$$\mu_p(\mathcal{F}) \leq \max_i \{\mu_p(\mathcal{A}_{i,t}^n)\}$$

■

We define for every  $t > 0$  and  $p < \frac{1}{2}$ , let  $a_{p,t} \stackrel{def}{=} \max_i (\mu_p(\mathcal{A}_{i,t}^n))$ . It remains to prove that for a fixed  $p < \frac{1}{2}$ ,  $\lim_{t \rightarrow \infty} a_{p,t} = 0$ .

A subset  $F \notin \mathcal{A}_i$  must intersect  $[1, 2i + t]$  on at most  $i + t - 1$  elements. If additionally  $F \in \mathcal{A}_{i+1}$  it must then contain both  $2i + t + 1$  and  $2i + t + 2$ . Thus,

$$\mu_p(\mathcal{A}_{i+1} \setminus \mathcal{A}_i) = \binom{2i+t}{i+t-1} \cdot p^{i+t-1} (1-p)^{i+1} \cdot p^2$$

Similarly,

$$\mu_p(\mathcal{A}_i \setminus \mathcal{A}_{i+1}) = \binom{2i+t}{i+t} \cdot p^{i+t} (1-p)^i \cdot (1-p)^2$$

Together,

$$\begin{aligned} \mu_p(\mathcal{A}_{i+1}) - \mu_p(\mathcal{A}_i) &= \mu_p(\mathcal{A}_{i+1} \setminus \mathcal{A}_i) - \mu_p(\mathcal{A}_i \setminus \mathcal{A}_{i+1}) \\ &= p^{i+t} (1-p)^{i+1} \binom{2i+t}{i+t-1} \left( p - (1-p) \frac{i+1}{i+t} \right) \end{aligned}$$

The sign of this difference is determined by  $p - (1-p) \frac{i+1}{i+t}$ . For a fixed  $p$ , this expression goes from positive to negative passing through zero once at  $i = \frac{(t+1)p-1}{1-2p}$ , and decreasing thereafter. When  $t$  increases, the maximum  $a_{p,t}$  is attained at larger and larger values of  $i$ . However, for increasing values of  $i$  (and any  $t$ ),  $\mu_p(\mathcal{A}_{i,t}^n)$  tends to zero since it is a tail of the binomial distribution with  $p < \frac{1}{2}$ .

■